DOCUMENT RESUME

ED 401 941 IR 056 216

AUTHOR Buterbaugh, Nancy Toth
TITLE A Database Evaluation Based on Information Needs of Academic Social Scientists.
PUB DATE May 96
NOTE 65p.; Master's Research Paper, Kent State University.
PUB TYPE Dissertations/Theses - Undetermined (040) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Access to Information; Comparative Analysis; Database Design; *Databases; Electronic Text; Evaluation Methods; Higher Education; *Information Needs; Library Research; Printed Materials; Reference Materials; Research Methodology; *Research Tools; Search Strategies; *Social Science Research; *User Needs (Information); User Satisfaction (Information); World Wide Web
IDENTIFIERS Barriers to Information; *Information Behavior Research

ABSTRACT

This study evaluates two databases, "Historical Abstracts" and REESWeb, to determine their effectiveness in supporting academic social science research. While many performance evaluations gather quantitative data from isolated query and response transactions, this study is a qualitative evaluation of the databases in the context of actual research. The evaluation tool was based on research done on the information goals and behaviors of academic social scientists. The information goals were divided into specific information tasks against which the databases were evaluated in the context of a simulated research project. The most notable project limitation was the linear design of the evaluation study itself. Although the databases were initially chosen for their differences (one is print, the other is electronic; one is an index, the other is a directory) the study indicates that both databases show strengths in the same task, that of extracting. General conclusions concerning databases designed to support scholarly research include: (1) in order for librarians to more efficiently support academic research it is important that they understand the stages of research; (2) every task should be matched to at least one tool--tools should be equipped to aid the academic researcher at each stage of his research; and (3) tools should be designed to allow maximum efficiency in research by performing multiple, related tasks, for example, extraction and selection. Appendices include overview questions, identification of key sources, REESWeb homepage, and lists of resources found in the second stage of evaluation of both databases. (Contains 32 references.) (Author/SWC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 401 941

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A database evaluation based on information needs of academic social scientists

A Master's Research Paper submitted to the
Kent State University School of Library and Information Science
in partial fulfillment of the requirements
for the degree Master of Library Science

by

Nancy Toth Buterbaugh

May, 1996

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Nancy J. Buterbaugh
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR056216

BEST COPY AVAILABLE

ABSTRACT

This study evaluates two databases, *Historical Abstracts* and REESWeb, based on their effectiveness in supporting academic social science research. While many performance evaluations gather quantitative data from isolated query and response transactions, this study is a qualitative evaluation of the databases in the context of actual research. The evaluation tool is based on research done on the information goals and behaviors of academic social scientists. These goals are divided into specific information tasks against which the databases can be evaluated in the context of a simulated research project.

This pilot study in this kind of evaluation yields informative results, but it will be even more effective after certain limitations which came to light during the course of this study are overcome. The most notable of these limitations is the linear design of the evaluation study itself. Although the databases were initially chosen for their differences (one is print, the other is electronic; one is an index, the other is a directory) the study indicates that both databases show strengths in the same task, that of extracting.

Master's Research Paper by

Nancy Toth Buterbaugh

B.A., University of Virginia, 1992

M.L.S., Kent State University, 1996

Approved by

Adviser Barbara H. Connell Date April 14, 1996

TABLE OF CONTENTS

Introduction and Literature Review 1

Methodology 7

Results 22

- Critique of evaluation study design 22
- Suggestions for improvement of the evaluation study design 24
- Suggestions for improvement of the data collection 25
- Summary of data 29
- Summary chart 42

Conclusions 43

- Suggestions for further research 48

Appendix A 50

Appendix B 53

Appendix C 54

Appendix D 55

Appendix E 56

Appendix F 57

Reference List 58

ACKNOWLEDGEMENTS

I am grateful for all the advice and encouragement from Tschera Harkness Connell, who would not let me settle for less than my best effort.

I am grateful to my husband who made supporting me and encouraging me in this project a priority even in the midst of his busy schedule.

I am grateful to Slavica Publishers, Inc. and to Chuck Gribble for supporting me and giving me the flexibility I needed to complete this project.

Introduction and Literature Review

As we enter the so-called "Information Age" there is a greater array than ever of tools at our disposal to gather information. Electronic bibliographic databases and the World Wide Web compete with traditional print tools, such as indexes and directories. What criteria guide the development of these new information sources? Are they being built upon the same assumptions concerning information needs and uses as the traditional print tools? Is there good evidence that these traditional tools really adequately meet the needs of users? It was my goal to explore in this study the utility of databases as relates to one particular set of users: academic researchers. For the purposes of this study the definition of database is that of "environments composed of people, equipment, and procedures organized to achieve specific information objectives." (Debons and Montgomery 1974). This definition is broad enough to include both print and electronic resources. For the purposes of this paper the terms "database" and "information retrieval system" will be used interchangeably.

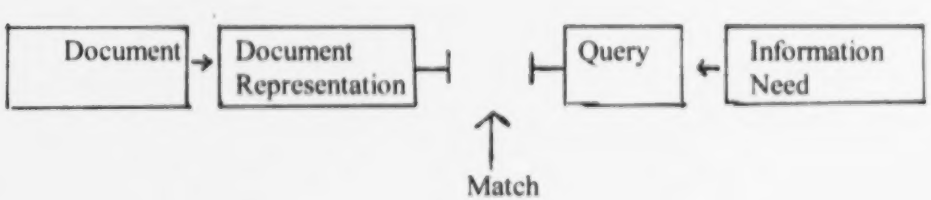
In spite of optimistic expectations that the access of online databases would forever change the face of research, evidence seems to indicate that their impact has been minimal in the work of researchers. Nor did their print predecessors receive heavy use from academic scholars. Stoan (1991) reviews several different bodies of literature which discuss the information uses of academic researchers. The literature generated by librarians tends to focus on measuring the frequency of use of various information sources by faculty members. Other research, which is performed largely by those outside the library community, focuses on scholarly communication as a whole, including "informal" channels. Both of these bodies of literature present strong indication that the use of bibliographic retrieval systems, in their traditional print form or their newer electronic form, by academic researchers is sparse. Studies on scholarly communication as a whole further indicate that scholars rely heavily on "informal" sources of communication, such as personal communication with colleagues, rather than on libraries, librarians and bibliographic tools. An example of well documented research which demonstrates these general tendencies can be found in Garvey. His research focused on gathering information from scientists concerning the information sources used to successfully complete the most recently completed stage in their research progression. For social scientists the largest percentage of researchers reported having

satisfied their most recent information need related to research from local colleagues or students (1979, 270). More recently, studies surveyed by Watson-Boone on the topic of research needs of humanities scholars also demonstrate that "these researchers have limited need for developing or using general bibliographic tools and various other secondary information services. In addition to tapping colleagues, they tend to look within initial print sources for ... references..." (1994, 213). At the minimum Stoan points out that "it can simply be stated that informal techniques for keeping up with the literature and retrieving materials useful for research prove to be satisfactory to scholars" (1991, 247). He further suggests that this system of scholarly communication has evolved because it "offers the highest reward-cost ratio" (1991, p. 248) for researchers. In making the most of their limited resources, academic researchers turn to the sources of information which are best suited to meet their needs. Stoan states that this approach "emphasizes information-retrieval channels that offer guidance from other experts in their fields." (1991, 248). If Stoan's assessment of the situation is accurate then there are significant implications for the field of librarianship. If an informal communication network among colleagues is the system that best satisfies the needs of academic researchers, then it is clear that tools and systems for information retrieval designed by information professionals have not succeeded in meeting the information needs of this community. A summary of research in the fields of database evaluation and of user needs assessments will indicate that there is little evidence that information tools have been created with the needs of academic researchers in mind.

Ellis (1992) gives a lucid and concise summary of the history of information retrieval (IR) evaluation. Information retrieval research began in the early 50s. In 1957, with the Cranfield experiments, the first paradigm for IR evaluation was established: the physical paradigm. It is based on the assumption that a physical system is an appropriate analogy from which to draw evaluation techniques. Characteristics of research done within the physical paradigm are quantitative evaluations performed in a "laboratory" environment. In other words, rather than testing the system while it is in normal operational use, a specific situation is designed for testing in which extraneous variables can be carefully controlled and evaluation is focused on isolated variables. Typical evaluation criteria for these experiments are the precision and recall of a retrieval set in a bibliographic database.

The second paradigm which Ellis discusses is the cognitive paradigm. This began in the late 1970s. Some of the earliest researchers associated with this paradigm are Oddy and Belkin. The assumption in the cognitive paradigm is that a useful information retrieval system will be based on an accurate model of the way in which the user's mind processes and categorizes information. Although the number of researchers in the cognitive paradigm is smaller than that in the physical paradigm, it appears that the former is increasing in influence as many researchers become convinced that the physical paradigm is insufficient. However, the cognitive paradigm leads to difficulties due to the sheer complexity of the human mind. To really build an abstract model which would adequately portray the workings of the mind in an information retrieval situation is a daunting task. Ellis concludes that neither of these paradigms will ultimately be able to produce an accurate theoretical model to inform IR research. There is an inherent duality in the nature of databases which seek to link physical objects, such as documents, with human users. While the physical paradigm is too simplistic to really deal with the subtleties of the task, the cognitive model, which seeks to remedy this fault, quickly becomes too complex to be useful. (Ellis 1992, 59-60)

Not only are the paradigms of information retrieval research inadequate, but the same is true of the model of information retrieval upon which this research is based. Bates refers to this as the "classic model" of information retrieval. It can be represented as follows:



(Bates 1989, 408)

The focus of this model is upon matching the user's query with the documents represented in the database and returning an answer, either in factual information or in the form of bibliographic citations, which best matches this query. This model of information retrieval is "the explicit statement of the implicit assumptions concerning the retrieval situation embodied in the laboratory

tests." (Ellis 1984, 269). Although followers of the cognitive school have criticized this model because it necessitates the user "translating" his or her query into the language of the system, rather than matching the document and question representation to the user's "mental map", they have not changed the fundamental assumptions concerning the components of an information retrieval process (Bates 1989, 409; Ellis 1993, 472) The inadequacy of this model can be attributed at least in part to the fact that its development has not been grounded in any sort of empirical research involving users and the actual tasks which they perform (Ellis 1984, 269). The assumptions about information retrieval made in the classic model are related to assumptions concerning the behavior of the user. These assumptions are summarized by Roberts (1982) in his description of an abstract "information man", akin to the "economic man" in early studies in the field of economics. He points out that the information man is a static and simplistic model, lacking any real consideration of complexities of human behavior. Nevertheless, this information man is assumed as the theoretical user in many information retrieval experiments. Given this state of affairs, the fact that the laboratory and cognitive models are unable to actually predict performance in operation situations is scarcely surprising.

It is at this point that the field of information retrieval evaluation is closely tied to information need and users studies. Ellis concludes his article with a call for research which would illuminate the information needs and behaviors of users for whom information retrieval systems are being developed.

If information retrieval systems evaluation can be traced back to the Cranfield experiments, the genesis of user studies can be found in the 1947 Royal Society Scientific Information Conference in the United Kingdom. Two papers were presented which became the seminal papers in the field of user studies, and which spurred much similar research. Dervin and Nilan (1986) identify a paradigm shift in user studies which began in the 1980s. This paradigm shift was the result of a realization that the current state of user studies lacked an adequate conceptual and theoretical basis to make meaningful predictions. The traditional paradigm which was in the process of being overthrown was primarily quantitative and objective. "It is one [paradigm] in which information is seen as objective and users are seen as input-output processors of information.... It does this [evaluates] by focusing on externally observable dimensions of

behavior and events." (Dervin and Nilan 1986, 16). This old paradigm was replaced, not by a single unified new paradigm, but by several different competing and overlapping paradigms. However, these paradigms have in common a general emphasis on more qualitative and subjective studies, with a greater respect for the human as a complex and central part of the information process. A strong influence on the new qualitative paradigm came from the field of social sciences, where, Patton (1986) describes an influx of more qualitative research in the 1970s. (Bawden 1990, 28-29)

Each new paradigm resulted from a critical response to a different flaw in the old paradigm. One target of criticism in the old paradigm is especially relevant to my study. Several articles in the early 1980s criticized the fact that user studies tend to be more oriented to the information system than to the user and his needs. In essence, studies have produced results which measure the use of a system rather than whether or not a users' needs were satisfied. Brittain points out that "[t]he majority of studies have been concerned with the need for and methods of finding out about and obtaining documents; few user studies have been concerned with the requirements for data or information as such" (1982, 143). In other words, the evaluation focused on the information in terms of how the system deals with it (as documents) not with how the user deals with it (as information for a specific purpose). Wilson further points out that "[s]uch studies may never address the central question of 'information need', that is, *why* the user decides to seek information..." (1981, 7). As summarized by Green "need is necessarily instrumental. That is, to need 'x' is always to do so with respect to some goal or end-state 'y'." (1990, 65) This is certainly true of information needs. A user usually does not need to see a copy of a particular article written by a particular author. What the user more often needs is information contained in that document. However, user studies have historically been guilty of ignoring this principle. "In the past a great deal of user studies research has suffered from a concentration on the *means* by which people discover information ... rather than upon the *ends* served by the information-seeking behaviour. (Wilson 1981, 10) This inversion of means and ends is a contributing factor to the poverty of theory in the field of user studies. Through this one-sided evaluation, research has failed to build up a model of what users are really looking for and what will adequately satisfy their information need. The studies simply reveal what users find, not what they may have really

needed. This ultimately results in 'marginality' (Wilson 1981, 12) of information services which are designed and adopted without consideration for the actual needs of the user.

Ellis, aware of the lack of an adequate model of the user for information retrieval systems, based his research on the grounded theory approach of Glaser and Strauss. This approach focuses on using empirical observation in order to build theories and models, rather than starting with *a priori* assumptions about the object of the study (Ellis 1993, 473). Ellis focused his study on the information needs of social scientists at the University of Sheffield. His aim was to acquire "micro-level information about the activities and perceptions of the academic social scientists which was considered necessary for detailed analysis of information retrieval system requirements" (Ellis 1989, 172). Ellis used a methodology of interviews with researchers concerning specific information needs and how they fulfilled these needs. The information from these interviews concerning their information needs were then analyzed to produce six major categories of information seeking behaviors which these social scientists employed. These tasks are: starting, chaining, browsing, differentiating, monitoring and extracting. The significance of this study is that is it avoids the means/ends inversion of so many user studies. Rather than simply noting what information systems or services the researchers used, Ellis took his study a step further to interviewing the researchers as to their *goals* in the information-seeking behavior. In describing these behaviors, Ellis outlines specific tasks which must be accomplished in order for the researcher to succeed in their goals in these behaviors. Bates uses Ellis's work and that of several others to propose a model of information retrieval that is fundamentally different from the classic model. She points out that observations of user's behavior with manual information sources indicates that the classic model of automated retrieval is incomplete in many ways. While the classic model presupposes a one-time well defined query, users typically follow a much less straightforward path. "Each new piece of information they encounter gives them new ideas and directions to follow and, consequently, a new conception of the query.... Furthermore, at each stage, with each different conception of the query, the user may identify useful information and references. In other words, the query is satisfied not by a single final retrieved set, but by a series of individual references and bit of information at each stage of the ever-modifying search." (Bates 1989, 409-410). Fundamentally, the classic model has isolated a single information transaction

from the larger context of the ill-structured and ongoing process of the user's information gathering. This more realistic model, which Bates calls "berrypicking" involves the behaviors documented by Ellis in his observations of social scientists. For this reason authors such as Bates and Ellis call for information systems to be designed specifically to perform the sorts of behaviors in which users typically engage to meet their information needs, rather than systems designed to meet the more static classic model. (Bates 1989; Ellis 1989).

The primary application for which Ellis intended his research is that of improved information retrieval system design. With a clear idea of what the goals of the user are, the system can be designed to be more effective and useful, thereby decreasing the marginality of the system. In his 1989 article Ellis also includes an overview evaluation of various information retrieval systems and the extent to which they allow the user to achieve the goals of the six different information seeking behaviors. However, the main focus of his article is to provide the basis necessary to design database. He does not attempt a systematic evaluation of any information retrieval system based on these information behaviors and goals. The aim of this study was to design an evaluation tool based on Ellis's study of information behaviors and to use this tool to evaluate two different databases on their utility in a simulation of social science research. My research question is: to what extent do databases (print or electronic) support academic researchers in performing the kinds of information behaviors which they typically use in the course of their work?

METHODOLOGY

The methodology which seems best suited to this study is a performance evaluation. Cronin defines this as "the process of systematically assessing effectiveness against a predetermined norm, standard, or expressed goal" (Cronin 1982). The usefulness of this method hinges on having well chosen criteria against which to measure effectiveness. These criteria must be both specific and meaningful by some objective standards. In other words, the possibility of setting performance goals which are prejudicially chosen to guarantee the "success" of the system must be avoided (Bawden 1990, 51). Swanson has a similar definition for what she terms

evaluation studies, which "investigate the degree to which stated goals or expectations have been achieved or the degree to which they can be achieved" (1978, 61).

Performance evaluations can be distinguished according to several different criteria. The performance measure can be in reference to *inputs*, that is the resources invested in the system; *outputs*, the extent to which the system fulfills its objectives; or *outcomes*, the effect which the outcome has on the user or the environment in general. (Bawden 1990, 50). Performance evaluations can be at a macro-level which usually involves gathering and comparing statistics that have to do with the overall information system. Micro-evaluations focus more on system performance in direct relation to a user. The most common topic for this kind of evaluation has been that of library reference services. The typical methodology has been some type of "unobtrusive" testing, in which the researcher pretends to be a user in order to evaluate the reference service. Reference services provide a convenient environment for this sort of testing: there is a well defined task to perform and judging the accuracy of the response is usually relatively straightforward. (Bawden 1990, 53, 141). This study is a micro-evaluation of an information retrieval system which focuses on outputs, that is the system's effectiveness in accomplishing certain objectives.

There are examples in the literature of performance evaluations used for information retrieval systems. However, these performance studies involving databases are based on the "classic" model of information retrieval, discussed earlier. These studies differ from my evaluation in several ways. These differences relate primarily to the context of the study and to the methods of collecting and reporting data. In most performance studies little, if any, consideration is given to the extent to which the tasks which the system performs in the course of the evaluation are synonymous with the tasks which are useful to the researcher in the context of his work. In many cases, the task for which the system's performance is evaluated is defined from within the system itself, not from the user. For example in the evaluation of CD-ROM encyclopedias the users were told to retrieve specific factual information known to be in the encyclopedia (e.g. Marchionini and Liebscher 1991). Even in Hert and Nilan's (1991) study in which the evaluators observed the users who came to use the online public access catalog in the course of their own research, there was an implicit assumption that the tasks which the catalog allowed them to perform were useful

for that research. The goal of the study was to minimize difficulties in performing those specific tasks by improved interface or design.

Hand-in-hand with this tendency to isolate the tasks which a database allows the researcher to perform from their larger research goal, is a similar tendency to evaluate each query-and-response interaction with the database in isolation from the rest of the researcher's interactions. This narrow and static evaluation is at the heart of Bates' (1989) criticism of the classic model of information retrieval. Evaluations are typically aimed at shedding light on the role of the cognitive abilities of the user (e.g. Ellis, Ford and Wood 1993), the interface of the system (e.g. Hert and Nilan 1991; Davis and Shaw 1989), or other factors which are essentially related to the interaction of the user and the system in the matching of a single query and response in the classic model. Therefore the data collected are usually quantitative data related to a query-response transactions. Examples include such factors as success at retrieving the specified information and retrieval time (e.g. Large et al. 1994; Marchionini and Liebscher 1991), search paths or commands used (e.g. Large et al. 1994; Cove and Walsh 1988; Hert and Nilan 1991), number of citations or usable citations (e.g. Rogers 1985) number of keystrokes to accomplish as task (e.g. Davis and Shaw 1989). Even the subjective questionnaires and interviews which are often used as a part of performance studies tend to focus on the user's evaluation of the system on an isolated task, such as the rated difficulty of finding a particular fact in the encyclopedia (Marchionini and Liebscher 1991, 42) rather than as part of a larger information need. These data, then, are so far removed from the actual research objective of the user, that it becomes difficult to use them as measures of the databases' utility in the context of those objectives.

My study was designed to reverse this trend towards isolation and to gather data which reflect the utility of the information retrieval system in the overall scheme of research. In this study, the evaluation began with the definition of a research objective, based on one of Ellis's interviews, which then formed the simulated context of the evaluation. Providing this context allowed the evaluation to proceed according to Bates' berrypicking model, thereby avoiding the static query-response model. The tasks against which the databases were evaluated had been defined *externally* to the databases and are based upon Ellis's actual observation of user's behavior and goals in the larger research context. The evaluation of the system consisted of

collecting data concerning the effectiveness with which the system allows tasks to be performed in the context of pursuing a research objective. Unlike a performance evaluation based on the classic model of information retrieval, this study has made every effort to make sure that the evaluation itself and the data collected are tied as closely as possible to a realistic research situation, in order that an evaluation of the system's effectiveness will be as meaningful for real-life users.

Another way in which this evaluation differs from many performance studies is that it makes no effort to quantify data. This decision was supported by several factors. First and foremost is that fact that the purpose of this study is to evaluate databases not primarily for the purpose of ranking or grading them on an objective scale. The origin of this study lay in the issue of the scarce use of databases by scholars and the reasons for this situation. Its purpose is to bring to light gaps in database performance vis à vis the user primarily in order to stimulate the design of systems which rectify this situation. Qualitative data are sufficient for this research objective. Other reasons for not using quantitative data are related to the difficulty of expressing the evaluation criteria in meaningful quantitative terms and are discussed later in this paper.

While this performance evaluation differs in several important ways from most of the library literature examples of performance evaluation, my methodology draws on the conceptual basis for performance evaluations from the field of librarianship. Another related source of inspiration for defining and refining my methodology, are the qualitative Beta tests that are used by software and information systems designers in order to perfect their designs before marketing a final product. These evaluations are designed to ensure that a system will be used once it is produced and therefore seek to ascertain that the systems is indeed performing tasks which will be useful to users in the context of their work. Many of these Beta tests include qualitative assessments of users' reactions to the utility of the system for their purposes (e.g. Woodsmall and Siegal 1988). In discussing the evaluation of man-machine systems, Rouse identifies three components of evaluation. Of these, the one with which this study concerns itself is *effectiveness* which he defines as "the extent to which use of the system leads to improved performance, makes a difficult task less difficult or enables accomplishing a task that could not otherwise be accomplished" (Rouse 1987, 596). In outlining his plan for evaluation he describes a first phase which includes "understanding users' activities" and then characterizing these needs in a formal

taxonomy of component tasks which support the main activities (Rouse 1987, 598). This is very similar to the evaluation tool which I used. An example of this kind of evaluation is described in a case study concerning the development and evaluation of a decision support system in which certain "measure of effectiveness" were identified in a simulation environment and then used as the basis for evaluation (Adelman and Donnell 1986, 291-2).

My evaluation measured the effectiveness of two different databases on performing tasks which had been identified as useful based on Ellis observation of the information gathering behaviors and goals of social scientists. Swanson, in developing a methodology for a performance study begins with determining values. A "value situation" is the result of "a subject (the judging individual), an object ... and a purposive relation between the subject and the object" (1978, 60). In other words, a value results from a context. In the case of this study, this context is provided by the research objectives of the social scientists interviewed by Ellis. Values are then used to judge criteria, which are "performance expectations or preferences in terms of which value judgements can be made" (Swanson 1978, 60). These criteria are the information tasks which the researchers expected to be able to perform in order for them to achieve their goals. Lastly, Swanson indicates that these criteria must be expressed in terms of objective measures, such as effectiveness. In Swanson's terms, the goal of this study is to evaluate databases in the context of academic research, on the basis of specific information tasks which researchers expect to perform, expressed in terms of effectiveness.

The research objective selected from among those presented in Ellis study is that of a scholar who is asked to design a course in an area of specialization within his field that does not correspond to the researcher's own area of expertise. If a database is defined as a system which is intended to achieve a specific information objective (see p. 1 above), it would obviously be meaningless to evaluate a database against a goal which it is not designed to support. It is true that there are no databases designed specifically to help an academic social scientist design a new course which is not in his area of expertise. However, there are many databases which have the more general objective of supporting academic social science research. It is not therefore unreasonable to evaluate these databases on their ability to support a specific task which falls within the broad range of the system's goal. The technique of selecting a specific case of a more

general phenomena is a common one in evaluative research, especially so in user-oriented evaluations (Bawden 94; 137-138).

The following is the text of the interview Ellis conducted with the social scientist engaged in designing a new course, followed by some of Ellis's comments:

> Last year I had to set up one [course] that was very much outside my normal range -- teaching the industrial revolution period as history ... I started with the last five years' *Economic History Reviews* [sic] which are standard, that is the economic historians' journal -- *Economic History Review*. I looked through that for the last five years what articles there were on that period, what books had been published on that period over the last five years, used those as a basis for a literature search. Read around there in journals and books, and from bibliographies in those to bibliographies in other things ... And then there are the standard specialist things like the Newcomen Society for Engineering History and the *Agricultural History Review* for agrarian history, and so on, and you go to these. It's the obvious ones to start with.
>
> This account includes starting (with a review type journal), extracting (employing the same journal), chaining, and differentiating (both general and specialist sources in the field).
>
> (Ellis 1989, 178)

The research objective of this social scientist was to find the information necessary to design this new course. In the course of pursuing this objective the social scientist interviewed by Ellis engaged in four specific information behaviors characteristic of social scientists: starting, extracting, chaining, and differentiating. In his study Ellis defines these tasks as follows:

> *Starting*: activities characteristic of the initial search for information
> *Extracting*: systematically working through a particular source to locate materials of interest
> *Chaining:* following chains of citations or other forms of referential connection between materials
> *Differentiating*: using differences between sources as filters on the nature and quality of the material examined (Ellis 1989, 178)

These four behaviors were divided into component tasks, the identification of which is based primarily on Ellis's descriptions of the researchers' goals in these behaviors. These tasks then formed the criteria (in Swanson's terminology) for this study. In order to simulate as accurately as possible the needs of a social science researcher with the objective of designing a new course, this

context was limited to the information behaviors which Ellis recorded and analyzed in his interview with the academic researcher performing this task. The evaluation was then carried out by the author for a variety of reasons. First of all the goal of this evaluation study is give a descriptive picture of the database's performance capabilities. The variations introduced by the different skills of users is not part of the scope of this study. Furthermore, the technique of using "surrogate judges" (Swanson 1978, 67) is a common practice in information science as well as other social sciences simply due to time and economic restraints in many studies. "Studies may seek an approximation of a situation. If surrogate data are understood and interpreted as estimates, ... the data can be indicative and perhaps sufficient for the study purpose" (p. 67). It is in this spirit in which the author served as a surrogate judge for the ability of a database to perform information functions necessary to an academic researcher in the preparation of a new course. A course was selected from the course description catalog of a major research university. The course selected, History 628, described as follows:

> **Revolutionary Russia 1880-1914**
> The revolutions of 1905 and 1917 from the standpoint of the revolutionary movements of the 19th century and changing social and political forces

This course was selected primarily because the author has a Bachelor's Degree in Russian Studies and therefore has enough subject knowledge of this field to be able to undertake a research objective in this field. While it is acknowledged that history is alternately considered humanities or social science depending on the criteria being used it is the opinion of the author that this course is satisfactory for the purposes of this evaluation for at least two reasons. First of all, the emphasis of the course according to the course description is "social and political forces", which are typical of the field of study of political science, which is clearly considered a social science field. The comment of an historian discussing specifically the issue of whether history falls in the realm of humanities or social science offered the following distinction between the traditional historian and the more recent field of social history which is characterized by methods and areas of inquiry typically associated with social science.:

> "Social history deals with the development of society, its social structure, economic factors and political influences on the society ... the other end, that is the traditional historians treat history as story telling. They do try to be objective, but it is generally a lot of biography" (Pandit 1992, 107)

The course description selected clearly deals with "social history," which can legitimately be considered a social science. A second defense for the choice to consider this course in the field of social science is that one of the social scientists interviewed by Ellis from the political science department described his research interests as "Russian history and politics" (Ellis 1989, 210). Since it is Ellis's study of social scientists upon which my evaluation criteria are based it seems reasonable to have performed my research evaluation in an area which overlaps with one of the social scientists interviewed by Ellis.

While it is true that many databases are designed to support academic research in general, and that the research objective of designing a new course falls into this broad category, it is also true that this objective necessitates the performance of several different information tasks. While a combination of databases should be able to support this research objective, it is not necessary that any one single database perform *all* the functions necessary to accomplish this objective. For this reason two very different databases, which are both well-known and reputable, were selected for evaluation. To my knowledge there is no database which has a scope as specific as political and social forces of change in revolutionary Russia. Therefore an attempt was made to choose broader databases, the scope of which could legitimately be expected to include historical and political information on revolutionary Russia. The first is the Internet resource developed by the University of Pittsburgh entitled REESweb (http://www.pitt.edu/~cjp/rees.html). Librarians and other researchers in the field consider this to be a premier resource. It is described as "A very comprehensive site, offering shortcuts to frequently used sources..." (Association of College and Research Libraries 1995, 40). Information on historical, political and social forces in revolutionary Russia would certainly seem to fall within the range of a "comprehensive" scholarly source on Russian and Eastern European Studies. The other database was the print form of Historical Abstracts. The scope of this source is defined as an index "to the world's periodical literature in history and the related social sciences and humanities". Again, although this source is

in the field of history it is obviously considered to span the boundary between social science and humanities. Within a database this broad it would certainly be reasonable to expect information on revolutionary Russia. The division between the two parts of historical abstracts (1914) falls right on the border of the scope of the new course. Although the title of the course is Russia 1880-1914, the course description mentions the 1917 revolution. For this reason it was necessary to use both parts of historical abstracts in order to adequately research the topic of the new course.

These two databases were deliberately selected to allow the greatest possible contrast, since the goal is not to compare the two, but rather to find sources which, in combination, would support the academic researcher's goal. The first contrast is that one source is electronic and the other is print. Secondly one is focused on Russian and East European information of all sorts, presumably including history, while the other focuses on history, which will presumably include Russian history. Also, REESweb is a directory resource, serving to point the way to other resources of many kinds, while Historical Abstracts is an index to the periodical literature of that field. Lastly, Historical Abstracts specifically targets bibliographical information while the REESweb includes information both on Russia directly as well as on literature written about Russia. This diversity between the databases is designed to cover a range of information sources available to social science researchers with the expectation that between the two resources all four of the information behaviors identified by Ellis in the process of designing a new course will be supported.

The information behaviors identified by Ellis which were used for this evaluation are: starting, chaining, differentiating and extracting. I have used his definitions and descriptions as a starting point, but in some cases I have added to them in order to define component tasks which are required in order to successfully complete the goal of the behavior.

Starting: activities characteristic of the initial search for information. This activity is made up of two component tasks:

Gaining an overview of a new subject area.

Identifying key sources (such as authors, papers, publishers, etc.) in the field. For the purposes of this study the identification of databases from which to begin research has already been performed.

Chaining: following chains of citations or other forms of referential connection between material. This activity includes three component tasks:

Forward chaining: finding sources which are based on the current source

Backward chaining: finding sources upon which the current source is based

Closure: the realization that the major papers relative to a study have been uncovered

Differentiating: using differences between sources as filters on the nature and quality of the material examined. The component tasks include:

Filtering by the substantive topic of the study

Filtering by perspective or school of thought

Filtering by quality or level of treatment

Extracting: systematically working through a particular source to locate material of interest. This activity includes:

Identification of potentially useful sources of information

Identification of relevant materials from within that source

(Ellis 1989, 178-187, 190-192, 198-200)

The following questions* are designed to collect data as to whether or not a database provides the information necessary to perform the information tasks listed above. These questions were identified primarily by drawing on Ellis's descriptions of the goals of the behavior which he observed. They are divided according to the task to which they are relevant.

---

*In the course of the evaluation it was necessary to rewrite several of these questions to avoid ambiguity. See pp. 26 for the revised questions actually used in the evaluation.

Gaining an overview.

Does the definition of the scope of the database or divisions within the database give any indications of the parameters of the field of study (i.e. dates included or excluded, mention of related topics which are *not* included in the database, etc.)?

Does the database indicate what subfields compose the area of interest? Indications could be a list of topics or categories included in the subject area, or descriptions of different parts of the subject area.

Does the database indicate relationships between subfields of the area of study? Examples could be hierarchical relationships, such as X includes Y or subdivision relationships such as this topic is divided into X and Y with a description of how X and Y differ from one another.

Identifying key sources.

Does the database indicate what authors, papers or publishers are considered important in the field? Examples could be any paper called foundational, seminal or something similar and/or any source which is frequently cited; scholars or authors considered to have "founded" an area of study or considered "experts" in an area of study, or publishers or journals considered central to a field.

Forward chaining:

Is it possible to gain access to the sources which have built upon a source of interest?

Backward chaining

Is it possible to gain access to sources upon which a source of interest is built?

Closure

Is it clear that the important sources in an area of study have been covered? Indications could be a repetition of the same already known sources in both forward or backward chaining, or some sort of summative bibliographic study which comprehensively surveys sources in an area of study.

Filtering by topic

Does the database allow the user to narrow his search to specific topics within a field? Examples could be through post-coordinate Boolean searching of keywords or descriptors, through the use of subjects headings, etc.

Filtering by perspective or school of thought

Does the database allow the user to narrow the search to only information from a certain perspective? Examples could be through subject headings which include information on perspective, or through narrowing the search to journals or publishers known to produce materials from a certain perspective.

Filtering by quality or level of treatment

Does the database allow the user to identify sources with a certain level of detail or by the recognition they have attained within the field? Examples could include again a filter by journal or publisher, as well as some sort of guide to indicate which sources are most widely used in the field or by which sources are most appropriate to what level of scholar.

Identification of potentially useful sources of information

Does the database bring together sources from which a scholar could extract useful information? Examples would be a run of a journal, or a publisher's catalog on a particular topic.

Identification of relevant material from these source

Does the database provide mechanisms by which the scholar can select relevant items? Examples would be abstracts, tables of contents, book "blurbs", subject headings, etc.

The above are indicators of the effectiveness with which each databases performs the component tasks of the research behaviors identified by Ellis. The evaluation of the effectiveness of the databases took place in two stages. The first stage evaluated the extent to which the information task described above can be performed using only the database. However, since most databases are secondary sources, a second stage of the evaluation included information which is found by following references from the database being evaluated to an additional information source, removed from the database by only *one* step. For instance, if the database gives a citation

to an article which gives an overview of the field, the database itself would not fulfill the task of giving an overview, but this task could be performed by using another source which was directly indicated by the database. It is useful to keep in mind that, depending on the information task, it is not always necessary to consult a primary source. For instance, in trying to gain an overview of the area of study, a menu of subject areas which shows hierarchical relationships might perform the overview function extremely well, even though it has not provided the scholar with any primary source. In the course of evaluating the database in this second stage, more than one additional source might be used, however the user must be referred to each additional source which is consulted directly by the database. In other words, this evaluation did *not* follow a reference from a source which was itself referred to from the original database. The second stage of evaluation evaluated whatever increase in the effectiveness of the database results from following references to an additional source.

While the boundaries of the print database are clearly defined by the bound volume, the boundaries of a database on the World Wide Web are much less so. For the purposes of this study a database was defined as the entirety of different pages and resources which share the same server name at the beginning of their address (referred to a URL, or Uniform Resource Locator). A URL with additional or different address specifications following the same server name was considered to be part of the same database. Only when a link connects to a completely different server was the user considered to have left the original database. For example when evaluating the resources of the REESweb page, any resources which began with http://www.pitt.edu were still considered part of the same database. While this definition is somewhat arbitrary, since there is no absolute standard on the names of URLs, but the in first part of the address (almost always ending in org, com, edu or gov) there is generally a recognizable name which is used consistently for all resources offered by the same server. It seems logical to consider all the resources offered by the same server to be part of the same tool, while the need to link to a different server is a level removed which could be considered roughly equivalent to putting down a bibliographic tool and picking up the journal article to which that tool pointed the user. At each stage of the evaluation the database was evaluated on the extent to which it allows the user to perform the information tasks listed by Ellis. First it was necessary to establish if the task could be performed at all. If the

task could be performed, several factors were considered in evaluating how effectively the database supports this task.

The first factor is the extent to which knowledge of the subject is supplied by the system rather than the user. Any one of these tasks can be performed primarily using the system or by primarily using the knowledge of the scholar. An example would be that of the various filters described in the behavior of *differentiation*. If the database allows the scholar to filter his search by the names of journals, then a researcher possessing the knowledge of which journals are known for their quality, could use that option to filter the search by the quality of the information. However, for a scholar who is unaware of what journal titles correspond to "high quality," the database does not supply that knowledge for him. An alternative would be a database which offered on option of filter information by "quality" which a scholar could use, even if he were unaware of the journal titles to which that filter corresponds. This factor was referred to as "expert knowledge required."

Another consideration in evaluating the databases' performance on these tasks is how straightforwardly the information is presented. In order to perform the task does the user need to gather information scattered all over the database and put it together to draw conclusions? For instance, the user could scan through many different citations and conclude that a paper which was cited very frequently was a "key source". Alternatively the database itself could present a list of "key sources" based on citation frequency. This factor was referred to as "straightforward presentation".

A final factor which was considered in all cases is the completeness with which the task is performed. An example would be if the database mentions a partial list of subfields which are included in the main field of study, but which do not necessarily comprehend the entirety of this field. This factor will be referred to as "comprehensiveness."[**]

These three factors were considered for each information task in the evaluation. In addition, as the situation warranted, other factors which seemed to play a significant role in the

---

[**] In the course of the evaluation it was determined necessary to clarify the definition of "comprehensiveness" and to add the term "precision." See pp. 25-26 for the details of these changes.

evaluation of effectiveness will be mentioned. This "additional comments" category is in keeping with the qualitative nature of this evaluation which expects to learn more about the databases as they are evaluated, rather than approaching the study with the assumption that all the important factors have already been considered.

In summary, the evaluation tool which was used evaluated the selected databases on how well they allow a researcher to perform four information behaviors identified by Ellis. These four behaviors were broken down into a total of ten component tasks. In describing these ten tasks I identified several questions, the answers to which indicate whether or not the database provides the information necessary to perform them. In the course of evaluating the databases each question was asked in order to determine whether or not the database provided this information. If it did, the performance was evaluated in light of expert knowledge required, straightforward presentation, comprehensiveness and any additional factors which the evaluator deemed noteworthy. In order to ensure standardization in the course of the evaluation a data collection sheet was used. Examples of this data collection sheet for the component task of gaining an overview are included in Appendix A. The other sheets followed an identical format. This sheet was filled in with the evaluator's comments concerning the tasks which can be performed by the database. In the second stage an additional copy of the same sheet was filled out. This sheet included comments which are limited to the tasks which can be performed by additional sources to which the evaluator was referred by the database. These sources could provide information useful for the tasks that was not supplied by the database, or that was not supplied as effectively by the database. The composite of the answers to the questions in each component task formed the evaluation for the task.

The evaluation results are presented in terms of a qualitative description of the performance of the databases on each task, taking into consideration the factors of expert knowledge required, straightforward presentation, comprehensiveness, and any additional factors. As mentioned earlier, no attempt has been made to quantify these results, and indeed there seems to be little reason to do so. The ability to perform one of the information tasks can not be regarded as a black-and-white quality. Even if a task could be performed, this performance was heavily qualified by the factors of expert knowledge required, comprehensiveness and

straightforward presentation. Furthermore, these factors themselves are expressed in terms of continua, rather than discrete quantities. The aim of this evaluation study was to build a tool which could give an overview picture of the usefulness of different databases for the process of academic research. This study can then serve as a basis for further research in this area and for refining the evaluation tool.

RESULTS

There were two stated goals in this evaluation. The first was to gain insight into how well the two databases to be evaluated support academic research. The second was to improve the evaluation tool designed for this study. Due to several difficulties with the evaluation tool, it is the feeling of the author that the most significant information gained from the study is that relating to the means of evaluation itself and the ways in which this evaluation could be improved. However, some significant insight was also gained in the evaluation of databases in general and of the two being evaluated in particular.

Critique of the evaluation study design

In order that the evaluation of the databases will be properly qualified by the difficulties in the study, I will begin with a critique of the evaluation method and tools which I designed as part of this research paper.

In this study on evaluation of databases, in order to define the project two initial resources were chosen from which to simulate academic research on the Russian revolution. The sources chosen were the REESWeb Home Page on the World Wide Web and *Historical Abstracts*. This initial limitation of sources, coupled with the linear progression outlined in my two stage evaluation plan which necessitated completing the evaluation of the database before evaluating the resources to which that database directed me, and then restricted the evaluation to only one source beyond the initial one, proved to be very limiting. Of the four information behaviors which I simulated in this study (starting, chaining, differentiating and extracting) both of the databases I chose to evaluate were primarily beneficial for the information behavior of extraction. The most difficult part of undertaking research in a new topic of study is starting -- the process of finding or choosing an entry point into the body of information on the topic. The behavior of extraction cannot be accomplished very effectively until the researcher has really at least partially

accomplished the behavior of starting. Since in the first stage of evaluation neither of the two tools I had chosen were very useful for starting, the process of extraction was hindered by the lack of either a "key" source upon which to build or a comprehensive list of the subtopics to research. In many cases, sources to which *Historical Abstracts* referred me could have been very useful in the process of starting, but I did not consult these until I had finished the first stage of the evaluation. Furthermore, in some cases the sources to which REESWeb and *Historical Abstracts* pointed were resources which might have performed the behavior of extraction better. However, since my study was limited to only following my initial sources to one following source I was forced to continue to use REESWeb and *Historical Abstracts* as the main source from which to extract without being able to use the resources to which they pointed.

Also, the description of the research performed by the professor whom Ellis interviewed was very limiting. Ellis identified that this professor engaged in starting, chaining, extracting and differentiating. These tasks, however, only encompass the initial stages of the research objective of designing a syllabus. Ellis's interview gives no information on how the information which the professor gathered actually was assimilated into the syllabus. The stated goal of this evaluation study was to recreate the aspects of the professor's research which Ellis described in the interview. However, in many cases the fact that my evaluation study was an incomplete simulation of the research objective made it difficult to decide what information was really relevant to this objective. This uncertainty raised some doubt as to the extent to which I was able to accurately simulate the first stages of a research project which I did not bring to completion. For instance, some of the resources identified in Stage 2 of the evaluation of *Historical Abstracts* did not contribute very well to the four information tasks of starting, chaining, differentiating and extracting. However, these same resources probably would have been very useful in actually writing the syllabus and would therefore represented a successfully completed process of extraction. Because I did not carry the research project through to completion, I could not be sure of this fact. Furthermore, since my simulation did not really complete the research objective of creating a syllabus, certain tasks which I had identified were not really relevant to this study. The information task of closure, for instance, is unlikely to be possible until the research is closer to completion than I was able to bring it in trying to model the research described by Ellis.

Suggestion for improvement to the evaluation study

All three of these difficulties, the two-stage limitation on the sources, the incomplete simulation of the research objective (creating a syllabus), and the fact that the resources I initially chose were, in the first stage of evaluation, unsuitable for the behavior of starting, have to do with the research context for this evaluation. They are the result primarily of the limitations of this particular study, rather than of the conceptual evaluation proposed here. Since this study was designed to be a pilot study in this kind of evaluation, the identification of these difficulties is an important contribution to refining the evaluation process.

The most obvious change which would greatly increase the authenticity of the research simulation would be to structure the evaluation in a less linear manner. My proposal was to go to each of the two databases in the evaluation, and test each of them to see how well they performed the four information behaviors of starting, chaining, differentiating and extracting in the context of creating a syllabus. Furthermore, the two stages of the evaluation were kept separate, the first stage was completed before the second one began. This structure implies a linear approach in which the series of questions was applied to each stage of each database in turn. A less structured approach would have allowed a better approximation of the research process. With this less structured approach the evaluator would pursue the research objective using whatever stage of the database and whatever information task is necessary at that point, rather than allowing the evaluation to be driven by the order of questions and stages of evaluation set up in advance. For instance, rather than attempting to extract from these two databases, with little or no overview information, I might have started with extracting likely sources for an overview, and then returned to extract more sources from my original database once I had gained an overview from these sources. In the case of *Historical Abstracts*, many of the sources found in the second stage were good for starting, and therefore would have greatly improved the process of extraction. This iterative process would have been more natural, and could have still included the use of the same data collection sheets. The difference would be that the order in which the different behaviors were evaluated would depend on the research, not on the structure imposed ahead of time.

Another method to increase the authenticity of an evaluation of this type, which would require a significantly larger time investment is for the evaluator to undertake a more full scale

academic research project with the intent of monitoring the usefulness of various databases in this goal. Rather than only imitating certain steps, as in this study, the evaluator would actually complete the research objective.

Suggestions for improvement of the data collection

The actual process of evaluation was highly effective. The behaviors described by Ellis comprehensively covered all of my "instinctive" ways of sorting through information. This confirmed that his categories might indeed provide an excellent tool for evaluating databases as they are actually used in academic research.

However, there were several difficulties with the data collection sheets which I had designed. For each question I had only one set of blanks concerning how, or if, the relevant component task was performed. However, in many cases the same task was performed in several completely different ways by the same database. This necessitated numbering several different responses in each blank to show that they related to the same method of performing the task. In the future I would design a sheet with a whole series of smaller blanks under each question in which I could describe several different ways in which the task could be performed. I would also include a space on the sheet next to each method to mark exactly what database record or source contained that particular example of completing the task. In order to be able to trace each method of completing a task back to the source in which I observed this method, I used a system of numbers coordinated to a separate sheet listing the sources (see Appendix B for a copy of one of the completed data evaluation sheets).Appendix C is an sample sheet of how I would redesign the data collection sheets in order to repeat this study.

There were a few terms used in data collection which proved to be ambiguous. In evaluating *Historical Abstracts*, I discovered some confusion in trying to evaluate comprehensiveness, which could be interpreted in at least two ways. Information could be comprehensive within a defined scope related to a task, like the date headings for the abstracts which gave complete information on the dates which were covered by each source. This definition was more similar to the way in which comprehensiveness was intended in my proposal. However, the abstracts offered more information than the date headings. This information differed in the fact that there was no predefined scope against which the abstract could be measured for

comprehensiveness. The more complete information allowed a greater precision of selection. To avoid confusion, I will continue to refer to comprehensiveness as the consistency with which a particular method provides the information required by a task, such as the date headings, which consistently give information on the time period for the purpose of selection. *Precision* will refer to the extent to which a method gives enough information to allow a task to be done precisely.

The questions concerning differentiation (filtering by topic, quality, etc.) were ambiguous. In Ellis's description of differentiating he focuses on examples of the scholar distinguishing between whole categories of resources based on the substantive topic, quality or perspective of their common source. For example, a scholar would select all the articles from a particular journal as potentially useful, based on the fact that that journal corresponded to the topic, perspective or quality which was relevant to his research objective. However, as worded in the original data collection sheets, differentiation could have been understood as filtering between single resources on an individual basis, rather than on the basis of their belonging to a group which could be distinguished by topic, quality or perspective. In taking data, in spite of the ambiguous wording of my data collection sheets, I limited differentiation to distinctions based on membership is a certain class of resources, rather than on differences between individual sources (see Q8, Q9 and Q10 below). The behavior of extraction, on the other hand, while using many of the same criteria (such as subject, type or quality) focuses on discriminating among the actual items on an individual basis (see Q11 and Q12 below). For future studies, the questions for differentiating should be rewritten to focus exclusively on filtering something based on its membership is class of items (articles in a journal, books published by a certain publishing house, etc.) not to distinguish between individual items. Also, the task of Filtering by quality or level of treatment, should be amended to Filtering by quality, level *or type* of treatment. In some cases I found myself excluding certain sources because they were of a particular type, like conference proceedings. This criteria was not explicitly stated on my data collection sheets, but was mentioned by Ellis. I therefore decided to perform my data collection as if this criteria were explicitly listed.

Lastly, rather than re-starting the numbering of the questions for each information task, I would have numbered them in one straight sequence. With the separate numbering for each task, in order to verify that I had evaluated a source against each question, I had to check that I had

covered each information task. It would have been easier to verify that I had finished an evaluation if I had been able to simply check that I had answered a certain number of questions, rather than having to remember each of the ten component tasks, and then remember which of these had more than one question associated with them.

Below is a revised version of the evaluation tool questions. In the discussion of the data I will use the numbering and terminology from this revised list of questions.

Gaining an overview.

Q1 Does the definition of the scope of the database or divisions within the database give any indications of the parameters of the field of study (i.e. dates included or excluded, mention of related topics which are *not* included in the database, etc.)?

Q2 Does the database indicate what subfields compose the area of interest? Indications could be a list of topics or categories included in the subject area, or descriptions of different parts of the subject area.

Q3 Does the database indicate relationships between subfields of the area of study? Examples could be hierarchical relationships, such as X includes Y or subdivision relationships such as this topic is divided into X and Y with a description of how X and Y differ from one another.

Identifying key sources.

Q4 Does the database indicate what authors, papers or publishers are considered important in the field? Examples could be any paper called foundational, seminal or something similar and/or any source which is frequently cited; scholars or authors considered to have "founded" an area of study or considered "experts" in an area of study, or publishers or journals considered central to a field.

Forward chaining:

Q5 Is it possible to gain access to sources which have built upon a source of interest to the research?

Backward chaining

Q6 Is it possible to gain access to the sources upon which a source of interest to the researcher is built?

Closure

Q7 Is it clear that the important sources in an area of study have been covered? Indications could be a repetition of the same already known sources in both forward or backward chaining, or some sort of summative bibliographic study which comprehensively surveys sources in an area of study.

Filtering by topic

Q8 Does the database allow the user to identify items from sources which focus on a specific topic? Examples could be a filter, either manual or electronic by which the researcher can exclude or focus on articles from certain journals or books from certain publishers which are known to cover a particular topic.

Filtering by perspective or school of thought

Q9 Does the database allow the user to identify sources which focus on a certain perspective? Examples are equivalent to those given in Q8.

Filtering by quality or level or type of treatment

Q10 Does the database allow the user to identify sources with a certain level of detail or by the recognition they have attained within the field? Examples could include again a filter by journal or publisher, as well as indications if an item belongs to a certain distinct class of information sources, such as conference proceedings or personal archives.

Identification of potentially useful sources of information

Q11 Does the database bring together sources from which a scholar could extract useful information? Example would a run of a journal, or a publisher's catalog on a particular topic.

Identification of relevant material from these source

Q12 Does the database provide mechanisms by which the scholar can select relevant items? Examples would be abstracts, tables of contents, book "blurbs", subject headings, etc.

Summary of data

As mentioned earlier, the validity of the evaluation of the databases is qualified by the limitations of this study. However, some informative conclusions can be drawn from observations of these databases. The actual results of the evaluation revealed that most of the information tasks could be performed at least some of the time in each database. However, these tasks were often completed in ways that were not straightforward. As expected, the fact that the databases *could* perform certain tasks was not really an indication that it was designed in a way that really facilitated the task.

REESWeb -- Stage 1. The REESWeb page itself is the most limited of all the resources in the tasks which it allowed. REESWeb does not give any means for gaining an overview, for chaining or for identifying key sources. A collection of resources ranging from information on gay life in Moscow to an exhibit on the artifacts of Genghis Khan is presented in paragraph form divided into rough categories such as "History, Geography and Sociology". It is possible to begin to recognize a certain set of references to which links appeared under several different categories. However, I determined that this did not amount to an identification of a "key source" (Q4). There is no claim by the REESWeb to select among the great variety of resources available on the World Wide Web, rather it is a comprehensive resource. Since the REESWeb is a directory of resources the fact that the same resources were listed several times on the REESWeb only indicates that the compiler of this directory knew of these resources and considered them relevant to several categories of research. It does not mean that there is any consensus among scholars that

these sources are important. Therefore, there does not seem to be any mechanism for identifying key sources in REESWeb.

The only information behavior to which the REESWeb does lend itself is that of extracting and, to a very limited extent, differentiation. This home page is a directory of resources on the World Wide Web of interest to Slavists. It therefore fulfills the first information task in extraction by identifying potentially useful sources of information (Q11). The user can then select sources from this directory. The REESWeb offers some help in identifying which resources would be most useful (Q12). Many of the sources to which the page offers links are given a brief description, such as "the Library of Congress's revelations of the Soviet Archives contains over 30 directories with documents covering topics such as the Cold War, Collectivization, Chernobyl, Stalin, the KGB and others". While these aids to extracting relevant resources are presented in a straightforward way, in the paragraph listing the resources, it is not a very complete method of helping the user to extract. Many of the resources are not described at all in the paragraph while in other cases the name or description would only be of help to a user who was already familiar with a resource such as the *Former Soviet Union Reading Room at the University of Kansas History Net Archives*.

Differentiation is possible in some cases in which either the name or the description of a resource, would indicate to the user that this resource is a particular kind of source, such as an archive of Frequently Asked Questions to a discussion group (Q10). A user with expert knowledge of types of Internet resources could filter out these kinds of resources based on the level of detail in which they are likely to treat a topic and by the type of "hit-or-miss" coverage. However this sort of filtering is only possible in a few cases and with a fair amount of familiarity with Internet resources.

In the first stage of the study, REESWeb does not appear to be a great asset to a scholar attempting to do research. The main behavior which it does allow is that of extraction. What information is given to aid in extraction is inconsistent and incomplete. Appendix D includes a printout of the first part of the REESWeb page for "History, Geography and Sociology".

REESWeb -- Stage 2. Since REESWeb is a directory resource it is to be expected that the bulk of its usefulness to scholars will be found in the second stage of the study, in which links are

followed to resources to which REESWeb has links. Indeed in the second stage REESWeb supports more information tasks than in the first stage. The URLs of the resources which I investigated in the second stage of the REESWeb evaluation are listed in Appendix E.

The Bucknell University Russian Program Russian materials page provides a chronology of the revolutionary period which was useful to gaining an overview of the subject. A chronological list of events which are considered significant in the revolutionary period gives information on subfields which compose a topic of study (Q2). It is provided in a straightforward way and requires no real expert knowledge. The resource implies comprehensiveness in the description "...major events from the beginning of Russian History". However, it appears that it makes no references to events (such as the French Revolution) which occurred outside Russia and which are considered to be significant to eventual outcome of the Russian revolution.

Another site archives the journal *Slavic Review* from Fall 1994 to the present, with the full text of articles. This archive includes several articles on specific events of the revolutionary period, such as the Lena massacre, which is considered to be significant to the revolution. However, I determined that these did not really fulfill the information task of gaining an overview since they did not make any attempt to indicate other subfields in the topic of revolutionary Russia but focused exclusively on one aspect of the topic. The database itself does not offer any background information, nor are relationships between subfields described (Q1, Q3).

There are instances in which it is possible to gain information on key sources in a field (Q4). This occurred for example in the text of an article in the *Slavic Review* which states "Historians of this period ..." and proceeds to name scholars who are presumably considered knowledgeable of this period. Although this does supply information concerning key authors in the field and it does not require expert knowledge, it is not presented in a straightforward way. The only way to gain this information is to skim through the text of articles in hopes that such a phrase will appear. Furthermore, no attempt is made anywhere to provide a complete list of "experts" so the information is not very comprehensive at all.

A resource which provides a mechanism for differentiation is the Association of American University Presses Online Catalog/Bookstore. If the user possesses enough expert knowledge to choose certain university presses which have the desired quality or perspective (Q9, Q10) he or

she can use this feature search feature as a way to filter based on these criteria. Also in some cases where the publishers' catalogs include abstracts it is possible to filter by type of treatment (Q10). For instance, if a work is listed as containing conference proceedings, someone knowledgeable of scholarly publishing would know that these are likely to include articles which treat aspects of a topic in great detail and are less likely to include the introductory or overview material necessary for starting in a new field, or for teaching a survey course. The actual searches in this catalog, which allow differentiation, are technically disqualified from the boundaries of this study because the URL of the page from which one can search the publishers' catalog is different from the resource to which REESWeb can immediately direct a user. The REESWeb provides a link to the resource http://aaup.princeton.edu. Once the user selects a search from this screen, the search is actually performed from the URL http://aaup.pupress.princeton.edu. According to the guidelines set in the proposal this constituted a second step from the original resource and is therefore disqualified. However, since both resources are provided by Princeton and the unconventional step of modifying the beginning segment of the address was taken I believed this resource to still be within the logical boundaries of this study.

In no cases is there an example of chaining, other than the conventional citations at the end of the articles in the *Slavic Review* which allow backward chaining (Q6) by examining the print sources upon which the article was based. The environment of hypertext seems eminently suitable to quick and easy links for either forward or backward chaining, and yet in the subject area I examined there were no examples of this. The links appear to be used primarily for connecting groups of related papers or for hierarchical connections between resources with a broader organizational scope (such as a university) and a narrower scope (the Slavic department). The lack of chaining mechanisms on the World Wide Web may be a symptom of the novelty of this medium and the fact that it is unlikely that much research has been done yet which would cite a World Wide Web address as one of its sources. Once such research begins to appear on the Web it is hopeful that chaining would be able to occur through hyperlinks.

The vast majority of the resources which I encountered on the World Wide Web provide an opportunity for identifying potentially useful sources of information (Q11). These include the Russian Revolution page as well as several publisher's catalogs and the text of the journal *Slavic*

Review. While the publishers' catalogs include a variety of topics other than the Russian revolution, they still identified a potentially useful subset of resources since they were scholarly presses and university presses.

These sources of potentially useful information vary in the extent to which they support the researcher's informed decision on extracting particularly relevant resources (Q12). In some cases, the entries for the books listed in the American Association of University Presses catalog contain only bibliographical information. In these cases, the user can extract based on the author or publisher if he or she possesses enough expert knowledge to chose certain ones among them. In other cases, such as Eastview Publication, as well as some of the entries in the AAUP Catalog, there are abstracts included which give more information to aid in extraction, such as information on the subject matter. There is little useful information on the Russian Revolution page for extracting useful sources. The sources were simply listed by names which are only partially descriptive in most cases, such as *Quotes from the Russian Revolution, The Russian Revolution in dates* or *Picture of Victor Chernov*.

The article titles in the table of contents of the *Slavic Review* provides a means of selecting which articles are of interest. The presentation is quite straightforward, just like a table of contents in the print form of a journal. In some cases a certain amount of expert knowledge is required in order to know that a particular article related to the Russian revolution (such as *The 9th circle: The Lena Goldfield Workers and the Massacre of 4 April 1912*). Furthermore, the title is not always descriptive of the contents, such as *What's in a name? Of Dog-killer Jews and Rasputin,* which is a discussion of regulations in tsarist Russia governing the rights of a citizen to change his family name. In general, using titles for extraction is imprecise since the title is not a comprehensive representation of the information in the article.

Probably the best mechanism for extraction, when present, are the abstracts in the publishers catalogs, such as the Eastview Publications catalogs, as well as many of entries in the American Association of University Presses Catalog. Also, most of the publishers catalogs, including the Association of American University Presses, offer Boolean searching of subject descriptors.

Summary of REESWeb. In all, the REESWeb page, including both stages of information retrieval, offers modest resources to cope with the information tasks necessary for an academic to begin research in a new field. Chaining is almost entirely absent from this resource. Gaining an overview is possible in a few cases, but is not comprehensive in any of them. With a fair amount of expert knowledge, differentiation is also possible in a the Association of American University Presses and Eastview Publications. In general I think it is safe to assume that the resources on the Web which I located from the REESWeb do not seem to be designed to really support the functions of gaining an overview, chaining or differentiating.

On the contrary it does appear that some of them, such as the REESWeb itself, the Association of American University Presses catalog, the Eastview Publications page and the *Slavic Review* are relatively well equipped for the process of extraction. In each case they fulfill the task of bringing together a set of potentially useful resources. Although, none of them has completely comprehensive means for selecting among the resources, in each case there was at least some information which, even without a lot of expert knowledge, would allow even a non-expert user to identify relevant sources. The main shortcoming in the area of extraction is the lack of systematic and comprehensive information. The titles in the *Slavic Review* are not always informative, descriptions of resources on the REESWeb are not always present, and, when present, not always complete, and the AAUP page does not include abstracts for all the entries. Again, although extraction was possible on the Web, a resource designed specifically to support this kind of information behavior probably should include more consistent and straightforward mechanisms for identifying relevant material.

Historical Abstracts -- Stage 1. The database *Historical Abstracts*, although chosen for its differences from the REESWeb, actually showed a surprising similarity to it in terms of the information behaviors which it supported. Again the tasks which are best served by this resources are those which make up the behavior of extraction.

As before, some of the other tasks are supported only partially. In terms of gaining an overview, the answers to the different questions (Q1, Q2, Q3) present mixed results. The subject headings according to which the abstracts were arranged are helpful in dividing Russian history into meaningful historical periods (Q1). In the first volume, which covers history until 1914,

Russia is listed as the Russian Empire. In the second volume it is listed as the USSR and Successor States. This history is then divided into periods by years. These dates and bits of geographical information helped to establish a background based upon which tasks such as extracting and differentiation will become more efficient. Many of the abstracts mention one historical factor or another, such as feudalism or the industrial working conditions, as important in the revolution. However, these abstracts dealt with only one or several factors. In no instance is there an attempt to give a whole overview of all the factors which contributed (Q2). Likewise, there was no information in *Historical Abstracts* which really attempts to indicate relationships between different subfields (Q3).

While there was no mechanism which is specifically designed to give information about key sources in the field (Q4), scanning *Historical Abstracts* from one year of publication allows for the casual observation that certain journals, most notably *Voprosy Istorii* contain a large number of articles which related to the topic of the Russian revolution. Although *Historical Abstracts* draws from approximately 2,000 different journals, its abstracting is considered selective (Sheehy 1986). Therefore, the fact that *Historical Abstracts* contains many citations for articles on the Russian revolution from *Voprosy Istorii* suggests that this journal is established as a reputable one among historians.

The only chaining activity which is possible from *Historical Abstracts* is to chain backwards to the original resources which are summarized in the abstracts (Q6). Since this is precisely the task around which this database is designed this information is presented in a straightforward and comprehensive way and their is no expert knowledge required. *Historical Abstracts* is well designed for this task.

The other task for which *Historical Abstracts* is well equipped to allow is that of extracting. By definition the database has performed the first task in extraction: that of bringing together potentially useful sources (Q11). The scope of this resource is defined as abstracts from "all branches of world history, 1450 to present" The abstracts include articles from major historical journals, recently published books selected from book reviews, and doctoral dissertations. The exact boundaries of publication dates were not stated. The 1994 issue of *Historical Abstracts* includes citations as far back as the late 1980s. This is a wealth of resources

for the researcher. Furthermore, in some cases, the abstracts can give leads to even more sources of information. For instance, an abstract of a review article might include bibliographic information, and maybe even a brief description of the sources reviewed by the article. Although in the instances of this which I encountered in my evaluation the information was not complete enough to be very helpful in selecting relevant material. I selected two books from a list of books which were being reviewed, but in both cases, the books did not turn out to be appropriate. In one case the book focused on economic changes in the revolutionary period, in another case the focus of the book was the post revolutionary period. In each case the inclusion of a subtitle would have given me the information I needed to reject the book, but in neither case was the subtitle included.

From the set of potentially useful sources identified by *Historical Abstracts* relevant material can then be identified (Q12). There are at least four methods provided by *Historical Abstracts* to aid the researcher in selecting sources of interest. These different methods demonstrate the relationship of the various factors, according to which the behavior of extraction can be evaluated. In some cases, such as straightforward presentation and precision, these factors are in inverse proportion. Therefore, having more than one method, some of which rate higher in one factor than another, can improve the overall balance of the database.

The first tool for extraction from *Historical Abstracts* is the subject divisions within the database. *Historical Abstracts* divides history into subject categories (such as economic history) and into geographical categories. Within these broad categories there are divisions by broad time period. This method is very straightforward and precise in its presentation through the table of contents and the headings, and required no expert knowledge. However, the subject headings do not allow for very precise selection. For instance, in Part A of *Historical Abstracts* for 1994 there were a total of 13210 abstracts. Even using subject divisions, such as Russian History 1855-1914, I was only able to limit the number of possibly relevant entries to 943. There are obviously many historical works on this period of Russian history which do not focus specifically on the development of the revolution.

The next tool for selection is that of the date headings on each abstract which identify the particular years within the broad historical division on which that work is focused. These headings

are comprehensive and require only a little expert knowledge. For instance, in Part B of *Historical Abstracts* I could safely eliminate any works which focused on the years after 1920, knowing that the revolution was already completed by this time. Since each abstract has its own date heading this presentation required more labor on the part of the researcher, but also offers an increase in precision.

Once the entries are located which are of possible interest, the user can turn to the abstract for more information. The abstracts, which range in length from 50 to 200 words, are the most precise method of selecting. They offer information on the content, format or type of information which is presented in the item to which they referred. From my evaluation in Stage 2, it appears that the abstracts gave me comprehensive enough information to select relevant resources. However, I cannot be sure that I did not miss relevant articles because other abstracts were not comprehensive enough in their descriptions. For the interdisciplinary topic of my research this could be a crucial factor. Many abstracts describe an article which covered a specific topic, such as the working conditions in factories, in great detail. However, they do not indicate that the article makes a connection between the topic described and the Russian revolution. Since this connection was the focus of my research I did not include these articles in the resources I chose for Stage 2, although it is possible that I excluded some useful resources because the abstract did not give enough information. The usefulness of abstracts is also limited by the necessity of skimming through 100 or more words and by the fact that they are not completely comprehensive, since a few entries offer only bibliographic information, or a very brief abstract.

The other method of selection in *Historical Abstracts* is the index. This method allows a very straightforward presentation, but it requires much expert knowledge and lacks in precision. However, my observations of the index were limited, since it did not seem to match any of the needs I had in the project I was simulating. For instance, the entry on the Russian revolution, is narrowly focused on the specific years of the revolution, rather than the political and social climate leading up to it, which was the focus of my research. Furthermore, even within this restricted scope of the index entry are were many items which were not at all related to my topic, such as an article on funeral rites used during the years of the Russian revolution.

In addition to summarizing the content, many of the abstracts also made mention of the kind of source from which the item was derived, from which it was possible to gain information concerning the type or level of presentation. This falls into the behavior of differentiating (Q10). For instance, in some cases, the abstract mentions that a source is based on "personal archives", or that it is a record of conference proceedings. In general an article based on personal archives will contain a high level of detail, and therefore will probably not be suitable for a survey course. Also conference proceedings generally include articles which treat narrow topics at a high level of detail. Usually however, there does not seem to be an attempt to make sure that the articles as a whole constitute a complete treatment of a subject or that the articles are even connected to one another other than in the broadest sense. For this reason, they too, are probably not suitable for a survey course which is looking to give an overview of a large number of topics and to show their connections to each other. In the course of my research I therefore filtered these kinds of resources out based on their level of treatment. This process requires some expert knowledge, since one has to be familiar with different kinds of academic resources. The information on the source is not presented completely straightforwardly, since I had to skim the text of the abstract. Furthermore, there is no reason to assume that *every* resource which is a conference proceedings, or based on personal archives or some other kind of resource which tends to have a high level of detail, is identified as such in the abstract. Therefore, this information, while helpful when given, does not appear to be given comprehensively. It therefore seems unlikely that the database is really designed to support this task of filtering by level of treatment. In terms of differentiating sources based on topic or perspective (Q8, Q9) it is possible, although there are no tools to aid in this process. Since the abstracts mention the publisher of books and the journal from which articles are taken, someone with a lot of expertise in the field can distinguish among items based on their sources, but even with the expert knowledge this would require a tedious process of sifting through each abstract.

In summary, while it allows some starting and differentiating, *Historical Abstracts* in the first stage of evaluation appears primarily helpful for chaining and for extracting. Backward chaining is possible in a comprehensive and straightforward way. The combination of subject

divisions, date headings and abstracts allow a very efficient and extraction process through which the researcher could increasingly narrow his or her focus to the items of greatest interest.

Historical Abstracts -- Stage 2. In the second stage of evaluating *Historical Abstracts* the resources to which I was primarily referred were books and review articles. Most of the other articles for which I read abstracts appeared to be too narrowly focused to be of use in the interdisciplinary research which I was simulating. Although there were many different resources included in this stage of the evaluation, my results seemed to be relatively homogenous in terms of the kinds of research tasks which they supported. A list of the resources consulted in the second stage of this evaluation can be found in Appendix F.

Most of the books are helpful for the task of Gaining an overview. In general the table of contents presents a relatively complete list of the subfields which the author believes to be relevant to the development of the Russian revolution (Q2). In many cases, the table of contents includes dates which indicate the chronological relationships of these different factors. Subheadings within the chapters break the topic down further and show hierarchical relationships (Q3). This information is presented in a straightforward way and requires little expert knowledge.

Information on key sources in the field is provided by review articles (Q4). They list names of authors who seemed to figure prominently in an area of research. In many cases, the articles trace the development of a particular kind of research, and in the process indicate landmark articles, or authors who pioneered a field. Although these review articles indicate key sources and do not require any particular expert knowledge, this information did not claim to be comprehensive. It is certainly possible that important authors or studies are left out. Furthermore, each review article focuses on one topic within the larger study on the revolution and so there were many aspects of the Russian revolution for which I did not find review articles. This difficulty is, most likely, an inevitable part of interdisciplinary research.

Information on key sources can also be obtained by looking through endnotes and bibliographies for certain authors or works which are frequently repeated. Unlike the repetition of names on the REESWeb, this repetition can be considered evidence that these sources are key sources. By definition, if they are often cited by scholars in the field, they are key resources. Of course, the impressions arrived at by skimming bibliographies are unlikely to be comprehensive

and might not even be precise. On the whole, skimming references is not a very straightforward method for gathering information on key sources, at least not for a broad interdisciplinary topic.

Most of the resources I found offer information for backward chaining (Q6). The review articles generally had citations to the resources being reviewed, as well as footnotes for other resources mentioned. The books typically had bibliographies and endnotes. Assuming these works were written with regard to scholarly conventions, this method of being referred to sources upon which a work is based would be comprehensive, although there was no way to verify this. There were no opportunities for forward chaining (Q5).

Even though, as mentioned above, certain authors and sources were repeated, it is difficult to really ascertain that closure had been reached (Q7). I think this is due to several factors, which are mentioned in my assessment of difficulties in the design of the study. The first is the fact that I did not continue to follow citations, but stopped at the first source to which I was directed from *Historical Abstracts*. Second, is the fact that I did not really bring the research project of designing a syllabus to completion, but stopped after the first stages of research. Lastly, is the interdisciplinary nature of the study I chose. It will take much more research to see closure when many different areas of expertise are being tapped into.

As far as identifying other potentially useful sources of information, the review articles and the bibliographies are both helpful for this task (Q11). In the case of the review articles, the information is presented more straightforwardly. In most cases the citations are listed before the actual text of the article. The text of these articles then provides much more information about content, perspective and quality which can aid in the selection of items (Q12). None of the books which I found had annotated bibliographies, so, although they provided references to hundreds of potentially useful resources, it was very hard to select form among them. In some case, references to a source within the text can be helpful in selection, but this requires reading the entire book in order to identify more useful sources, which is not a very straightforward way to gain information on other possible sources. Neither the review articles, nor the book bibliographies make any explicit claim to be comprehensive. As with REESWeb, the full usefulness of these sources for extraction could not really be assessed because the limits of my study did not allow me to go on and extract further items from the sources to which I was referred my initial source. However, it

seems relatively clear that the review articles, in particular, are designed to support this function of extracting.

In terms of differentiating, none of the sources I found gave a lot of information about sources in general, such as journals or authors. Even the review articles are generally limited to the specific items being reviewed. In general, any kind of distinction based on topic, level or perspective appears to have to be done on an item-by-item basis, which does not meet the criteria for extraction (Q8, Q9, Q10). It is possible that with more expert knowledge, this would not have been the case. In several instances, once I actually had a book I could tell from the preface or other front matter that it was from a Marxist perspective. (Beaud (1993), for example, was dedicated to those who have "fought and still fight" for socialism). However, nothing I encountered in my research really seemed to equip me to differentiate between items based on their membership in a larger category.

Summary of Historical Abstracts. As a whole, *Historical Abstracts* appeared to be quite helpful in identifying useful sources of information and in aiding in extraction from them, which is, of course, the purpose for which this tool is designed. The sources to which *Historical Abstracts* pointed me performed virtually all the tasks for which I was evaluating. However, the main task which they seemed to perform well was that of giving an overview of the field. The books identify many of the subfields through their tables of contents. Review articles are especially helpful for identifying key sources. While backward chaining is certainly possible, in the books the number of sources identified is often too overwhelming to be useful. Review articles are more helpful in giving information about the sources upon which they are based. The review articles also are good sources for extraction, although for an interdisciplinary topic such as mine, it is unlikely that there would be any article or even collection of articles which would adequately cover the whole topic. Furthermore, as mentioned above, the second stage of *Historical Abstracts* does not present a complete evaluation of the potential of the database, because many of the sources located in the second stage probably supported functions not included in my simulation of this research project.

Summary chart

Below is a chart which summarizes the usefulness of the two databases at each stage in the evaluation process in response to each evaluation question. The ability of the database to perform the tasks are ranked according to the following scale:

+ The evaluator would choose this resource based on its efficiency as an aid in accomplishing the information task.

0 The evaluator could accomplish the information task using this resource, but would prefer to find a more efficient method for the task.

– The evaluator would not use this resource to accomplish the information task because it is either impossible, or extremely inefficient..

Behavior		REESWeb Stage 1	REESWeb Stage 2	*Historical Ab.* Stage 1	*Historical Ab.* Stage 2
Starting	**Q1**	–	–	0	0
	Q2	–	0	0	+
	Q3	–	–	0	+
	Q4	–	0	0	+
Chaining	**Q5**	–	–	–	–
	Q6	–	0	+	+
	Q7	–	–	–	–
Differentiating	**Q8**	–	–	0	–
	Q9	–	0	0	–
	Q10	0	0	0	–
Extracting	**Q11**	+	+	+	0
	Q12	0	0	+	+

Keep in mind that if the database receives a positive rating for one function which is a component of a certain behavior, and it receives a negative rating in the other questions which make up that behavior, the usefulness of the database for that entire behavior might be seriously compromised.

CONCLUSIONS

In order to be meaningful, this kind of evaluation must emulate the actual research done by academics. It is also important that the evaluator actually complete the research task which is being simulated in order the ensure that his or her actions in the initial stages are really consistent with the context of the research. In this way information scientists can evaluate tools in more and more realistic situations and design tools which are closely allied with the actual needs encountered by academics.

Another aspect of research which became evident through this study is the interconnectedness of the various tasks involved in completing a research project. In designing an evaluation study this factor must be addressed in two ways. In the first place, the linear design of my study proved to be limiting, by evaluating each database on each task in a rigid two stage process, following a checklist of tasks. Better results would have been obtained had a more natural, iterative research process been followed in which the information need of the moment would have dictated the task, rather than the checklist and two stage structure which was used. The other aspect of this interconnectedness which must be considered is the fact that in order to simulate a research project and evaluate the information sources used in it, it is likely that many sources will have to be used together, as different information tasks become necessary at different points in the research. The fact that I was trying to evaluate databases on the behavior of extraction when I did not have any way in the linear design of my study to have gained an overview, highlights the fact that effectiveness of one task is dependent on the effectiveness with which the prior task has been accomplished.

Since in most research projects, a single database will not be able to perform all the necessary information tasks, it will be necessary to use multiple databases in the course of a research simulation. To some extent, though, it will become difficult to separate the evaluations of the various tools, especially when dealing with secondary sources such as the databases which I chose. Many of the information tasks, such as chaining and extracting, by definition involve pointing the user to another source. It makes sense to include the quality of this next source in the evaluation of the initial source. However, as the levels of referral progress it is difficult to determine where the evaluation of one source ends and the next begins. Should the evaluation of a

source used for starting include the quality of the final product at which the researcher arrives? Obviously, if the scope of the evaluation becomes too broad, idiosyncracies of the researcher will be introduced and will affect the validity of the results. Any evaluation study will have to seek to take into consideration the interdependency of information tools within the context of a research process, while still narrowing the study enough to make meaningful observations. This interdependence of information tools is a continual reminder that these tools are not used in isolation, but rather as part of a larger research context. The lack or underdevelopment of tools to support even one of the many research tasks necessary to complete a project will result in a general diminishing of the quantity and quality of that project, as scholars are hindered in the successful accomplishment of one link in the interconnected chain of the research process.

As far as the actual evaluation of the databases is concerned, my conclusions are qualified by the acknowledged limitations of the study. However, there are still meaningful conclusions to be drawn. In the first place, it is interesting to note that, although I picked two resources which were different in many ways, for the specific purpose of providing resources to aid in as many of the information tasks as possible. In spite of this, both of my resources seemed to be primarily useful for the same task, that of extraction. So the common status of these two databases as secondary sources seems to give them more in common than the differences in their format and scope.

One observation that is sparked by my evaluation of REESWeb, is the question of whether all databases are designed around any specific task among the ones which Ellis describes in his article. As far as I can tell, the function of the REESWeb appears to be to pull together as many resources as possible on the World Wide Web which have to do with Russia and Eastern Europe. It certainly appears (although I did not evaluate this directly) that the REESWeb performs this function quite well. It is also very possible that this function is important for certain tasks for certain users. However, this sort of generic directory function does not really correspond to any of the functions which Ellis identified as having been part of the work done by the scholar who is designing a new syllabus. While it pulls together many resources from which the scholar could extract, it does not provide comprehensive enough information to aid the scholar in selection among these resources. Therefore, REESWeb did not prove very useful for the sort of research I

was simulating. As other kinds of research projects are simulated, the evaluation of REESWeb could prove to be dramatically more positive if it is evaluated in the context of a task which it was designed to support.

As far as the evaluation of *Historical Abstracts* is concerned, it was hampered by the linear design of my study, which resulted in using it for extraction without having an overview. It is ironic that some of the resources which I discovered from this extraction were excellent for gaining the overview which I needed to really do the extraction well. Had I followed a less structured evaluation, the information I gained from the sources to which I was referred could have been used to refine my extraction process. This is a perfect example of the iterative research process described by Bates (1989) in which information gathered at a later stage in the research results in a change in the original query. While some of this iteration is surely an inevitable part of the research process, my study raised the question of whether this "berrypicking" process could be streamlined by using resources that were well designed for each step of the research process. For instance, if resources had been available that gave a succinct and comprehensive overview of the field, could the starting stage have been mostly completed before extraction began, rather than having the two mingled together? The professor interviewed in Ellis' article started with a resource which was more focused on reviewing sources, which may have been better for gaining an overview, than the resources I chose. It would also be more helpful to know more precisely the order followed by the professor whom Ellis interviewed. Did the professor gain an overview from these reviews and then go on to extraction? Or did he engage in an iterative process of extracting good starting sources, and then going back to extraction after he had gained an overview?

From the limited confines of this study, it is not clear whether appropriate tools exist for each stage in this research project exist. Although I was able to gain some information on key sources from the review articles I read, I still did not gain a lot of overview information from them, since they tended to focus on more narrow topics than the one I had chosen.. Since part of this particular research objective of designing an interdisciplinary course is determining the topics to be covered, there is no single starting source that could perform this task. However, consulting several sources, such as textbooks or review articles which covered a broader range of topics could have given a relatively comprehensive basis from which to begin researching. In a different

kind of research, in which a professor is working in a field with which they are already familiar with the major subfields, and important journals and authors, the process of starting might come largely from their own personal background knowledge. In spite of the fact that starting tools will not always be able to give a complete overview of the exact aspects of a topic which the researcher needs, or the fact that in many cases a researcher will not need any outside sources for the task of starting, there is still much research which would benefit from tools which helped to develop an overview for the researcher before they began extracting.

Another behavior which *Historical Abstracts* did not support very well, but which could have been useful is the task of differentiation. A one year volume of *Historical Abstracts* included approximately 900 abstracts which could potentially relate to the Russian revolution according to the subject divisions. To skim through all of these for the purposes of extraction was a long process. Had I been able to eliminate some based on the quality, topic or level of treatment of the journal or other source, this would have significantly shortened the process.

The observations from the evaluations of these databases can give rise to certain general conclusions concerning databases designed to support scholarly research. First and foremost, in order for librarians to more efficiently support academic research it is important that they understand the stages of research. Tools should be created with the specific purpose of supporting tasks which are part of a academic research, not simply because the information is available. The REESWeb page is a poor use of resources if it does not support some identifiable information need. In other words, every tool should be matched to at least one task. The need for the purpose of a database to be clearly defined is further accentuated by the inverse relationship which appears to exist between certain factors inherent in different methods of performing certain tasks. As noted in the discussion of *Historical Abstracts*, the straightforward presentation of information is inversely proportional to its precision, and to some extent, the expert knowledge required. In cases such as these where it is not possible for a database to be maximally efficient in all areas, priority choices must be made between factors like expert knowledge and straightforward presentation. The only valid basis for making such choices is the research context in which a tool is intended to be used. If it is a tool for specialists in the field, then there is no

problem in making its presentation more effective by assuming a high level of expert knowledge. However, a tool to be used by beginners cannot follow the same procedure.

Secondly, every task should be matched to at least one tool. There must be tools which are equipped to aid the academic research at each stage of his research. The phenomena encountered in my evaluation of *Historical Abstracts*, in which one task (extracting) was hindered by the lack of the appropriate tool for a task which preceded (starting) results in a very tedious process. In fact, it might be the case that the "berrypicking" procedure through which research is conducted could be much more streamlined if researchers had access to tools which were effective for each stage of their research. To determine what, if any, gaps exist in tools supporting academic research should certainly be a high priority in research in this area.

Another lesson concerning databases which can be drawn from this evaluation is that careful consideration should be given to information tasks which are closely linked. When at all possible tools should be designed which allow maximum efficiency in research by performing related tasks. A relevant example is that of the REESWeb page. While this page performs the first stage in extraction, that of pulling together potentially useful resources, it does not provide adequate means for the second stage, that of selecting among these resources. These two tasks which together form the behavior of extraction cannot really be separated. In some cases, such REESWeb, the task of selection can still be accomplished with a little bit more effort: the researcher needs to follow each of the links to evaluate the resource directly. However, in other cases, such as that of the long bibliographies in the books without annotations, the resources are rendered virtually useless (for the particular behavior of extraction against which it is being evaluated) due to the lack of information given to aid in extraction from the list of potentially useful resources.

While the results of my research do not demonstrate that the field of librarianship has not provided the tools necessary to complete each stage of research in various different academic research objectives, it certainly leaves the question open. Further studies on how research is done, and on the effectiveness with which information tools can be made to support this research seems warranted. I do not believe that there is a lot of evidence that information tools are clearly designed around tasks which are known to be necessary for certain research projects, and are

designed with the context of such research projects in mind. Nor is it apparent that all common research tasks are well supported by databases designed for these purposes. In short, this study raises the question of whether the field of librarianship has really provided as well as is possible for the needs of academic researchers, and to provide the beginnings of the evaluation tools necessary to answer this question.

Suggestions for further research

In keeping with the general emphasis of this study on evaluating information tools within the context of actual usage, an important area into which continuing evaluation research could be directed is that of gathering more information from academic researchers about the start-to-finish process of completing a research objective. This information could help to ensure that the methods undertaken by evaluators would be a good approximation of those actually in use. It is possible that this research would lead to the conclusion that there is little similarity among academic researchers in how they actually go about eliminating certain information, while keeping other information, or how they choose which sources are worth pursuing. Even this realization would be useful to evaluators in order that they can make sure the scope of their simulations at least a rough approximation of the range of realistic scenarios.

A related goal is that of understanding how research methods and the need for certain tools vary according to the research objective. The research objective which I simulated is unique in several ways. First of all, the nature of beginning research in an unfamiliar area is very different from doing new studies in an area of expertise. The whole behavior of starting would become a much less significant part of the study for a scholar who was operating within a field in which he was already familiar with key sources and already possessed an overview. Another unique aspect of the research objective I chose is that the topic I was researching was broad and interdisciplinary. It included history, politics, economics and sociology. The topics which would be covered in a ten week course are much broader and multifaceted than the sorts of topics for which research articles are written. The sort of evaluation I performed should be repeated with varying research objectives. It may come to light that most databases are good for supporting certain kinds of research objectives, while there are relatively few tools to support other kinds of research.

This method of studying the process of academic research further, but still performing the evaluation by a process of simulation allows the evaluation to be done at the actual moment of the research, rather than through reconstructing the research after the fact and from a second party, the academic scholar, whose focus was on his research, not on evaluation. Studies of this kind would, of necessity, require large amounts of time and resources. However, this effort appears to be warranted by the situation described by Stoan (1991) in which academic researchers rely primarily on informal means of gathering information, while bibliographic tools, the creation of which also require a large amount of resources by librarians, remain vastly under used.

Appendix A

Overview Q1

Does the definition of the scope of the database or divisions within the database give any indications of the parameters of the field of study (i.e. dates included or excluded, mention of related topics which are *not* included in the database, etc.)?

provides this information	
expert knowledge required	
straightforward presentation	
comprehensiveness	
additional comments	

Overview Q2

Does the database indicate what subfields compose the area of interest? Indications could be a list of topics or categories included in the subject area, or descriptions of different parts of the subject area.

provides this information	
expert knowledge required	
straightforward presentation	
comprehensiveness	
additional comments	

Overview Q3

Does the database indicate relationships between subfields of the area of study? Examples could be hierarchical relationships, such as X includes Y or subdivision relationships such as this topic is divided into X and Y with a description of how X and Y differ from one another.

provides this information	
expert knowledge required	
straightforward presentation	
comprehensiveness	
additional comments	

(12) (2) HF

Appendix B

Identifying key sources Q1

Does the database indicate what authors, papers or publishers are considered important in the field? Examples could be any paper called foundational, seminal or something similar and/or any source which is frequently cited; scholars or authors considered to have "founded" an area of study or considered "experts" in an area of study, or publishers or journals considered central to a field.

(4) repetition of author in bib (11) Zeuinchavsky

provides this information ① ② ③	review article references cited authors "such as" (1) the word notes summarize origin of school of thought - some use "seminal" author's intro traces development of thought (3)
expert knowledge required ① ② ③ ④	assume they represent key names in some field What one? explains a lot knowledge of various contributors (Albert Camus) -
straightforward presentation ① ② ③ ④	embedded in text " note " " yes
comprehensiveness ① ② ③ ④	no pretense implies coverage of development of thought doesn't claim to be - implied not necessarily
additional comments	Source Review frequently cited (12) impressionistic review of a circles of authors published about the same time less citing each others... not a formal citation analysis indication of "in groups" research

BEST COPY AVAILABLE

Appendix C

Overview Q1

Does the definition of the database or divisions within the database give any indications of the parameters of the field of study (i.e. dates included or excluded, mention of related topics which are *not* included in the database, etc.)?

Method 1 _____

Source_____

expert knowledge	straightforward pres.	comprehensiveness	precision

Additional Comments:

Method 2 _____

Source_____

expert knowledge	straightforward pres.	comprehensiveness	precision

Additional Comments:

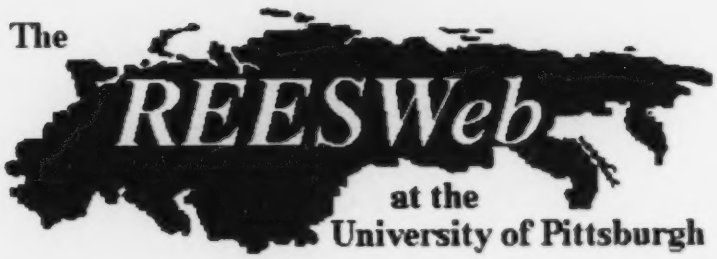
Method 3 _____

Source_____

expert knowledge	straightforward pres.	comprehensiveness	precision

Additional Comments:

Appendix D



REESWeb : Russian and East European Studies Internet Resources

History, Geography and Sociology Resources

This resource is sponsored and maintained by the University Center for Russian and East European Studies of the University of Pittsburgh

Click here to return to the Main REESWeb Page .

This page contains Russian, Soviet, and Former Soviet Union, Polish and General and East European Resources

---

Russian, Soviet and Former Soviet Union Resources

The Cybermuslim Information Collective's Interactive History Project features a 'collective hypermedia history' of Turks and Tatars. This exhibit is the first web endeavour of the IHP. It is hoped and intended that the portrait of Muslim Life in the 19th century Russian Empire constructed here will be the result of the collective efforts of its users. Contact Mas'ood Cajee about the project.

Alexander Artsyukhovich's Belarus Homepage provides sections on Belarusan history, Maps, Belarusan national anthem, symbol and flag, Belarusan Culture and Traditions, Attractions in Belarus, City guides, Advice for tourists, CIA 1994 annual report. Statistics. Governmen, Belarus on WWW, News on Belarus, and more.

The Center for Social Science Computation and Research at the University of Washington, has an on-line site for Russian and Soviet Archival Data. Some of these data files are available directly through the WWW, while others can be requested from tape.

The Black Hole Peripherals, Inc. home page contains a fascinating image collection, including: photos of old Russia, Russian money photo collection, old Russian commercials and advertisements 1880-1913, old menus of Russian restaurants, Russian leaflets of WWI, and photos of Moscow, August'91, Crimea,

Appendix E

Resources found in the second stage of evaluating REESWeb

American Association of University Presses Online Catalog/Bookstore	http://aaup.princeton.edu
Bucknell Russian Program Russian Materials	http://www.bucknell.edu/departments/russian/materials.html
Eastview Publications	http://www.eastview.com
The Russian Revolution	http://www.csv.warwich.ac.uk/~eshri/russ/rusrev.html
The Slavic Review	http://ccat.sas.upenn.edu/slavrev/slavrev.html

Appendix F

Resources found in second stage of evaluating *Historical Abtracts*:

Anan'ich, B.V. 1992. *Krizis vlasti i reformy v rossii na rubezhe XIX-XX vekov v issledovanniakh amerikanskix istorikov*. Otechestvennaia Istoriia 1992(2):208-211.

Beaud, Michael. 1993. *Socialism in the Crucible of History*. Atlantic Highlands, NJ: Humanities.

Davies, R.W., ed. 1991. *From Tsarism to the New Economic Policy: Continuity and Change in the Economy of the USSR*. Ithaca, NY: Cornell University Press.

Frankel, Edith, Frankel Jonathon and Knei-Paz, Baruch, eds. *Revolution in Russia: Reassessments of 1917*. Cambridge, England: Cambridge University Press.

Galili, Ziva. 1990. Workers' strikes and revolution. *International Labor and Working Class* (38): 68-80.

Moon, David. 1992. Agriculture and peasants, industry and workers, political parties and revolution: Recent books on Russian history. *European History Quarterly* 22(4):597-604.

Mosse, W. E. 1992. *Perestrioka Under the Tsars*. London and New York: I.B. Taurus.

Naarden, Bruno. 1992. *Socialist Europe and Revolutionary Russia: Perception and Prejudice 1848-1923*. Cambridge, England: Cambridge University Press.

Saunders, David. 1992. *Russia in the Age of Reaction and Reform 1801-1881*. London and New York: Longman.

Service, Robert. 1991. *Lenin: A Political Life Vol 2: Worlds in Collision*. Bloomington, IN: Indiana University Press.

Suhr, Gerald D. 1989. *1905 in St. Petersburg: Labor, Society and Revolution*. Stanford, CA: Stanford University Press.

Ward, Chris. 1990. *Russia's Cotton Workers and the New Economic Policy: Shop-floor Culture and State Policy 1921-1929*. Cambridge, England: Cambridge University Press.

Reference List

Adelman, Leonard, and Michael L. Donnell. 1986. Evaluating decision support systems: A general framework and case study. In *Microcomputer decision support systems: Design, implementation, and evaluation*, ed. Stephen J. Andriole, 285-309. Wellesley, Mass.: QED Information Sciences.

Association of College and Research Libraries. Slavic and East European Section. 1995. *SEES Newsletter*, no. 11.

Bates, Marcia J. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online Review* 13(5): 407-424.

Bawden, David. 1990. *User-oriented evaluation of information systems and services*. Brookfield, VT: Gower.

Beaud, Michael. 1993. *Socialism in the Crucible of History*. Atlantic Highlands, NJ: Humanities.

Brittain, J. M. 1982. Pitfalls of user research and some neglected areas. *Social Science Information Studies* 2: 139-48.

Cove, J.F., and B.C. Walsh. 1988. Online text retrieval via browsing. *Information Processing and Management* 24(1): 31-37.

Cronin, B. 1982. Performance measurement and information management. *Aslib Proceedings* 34: 227-36. Quoted in David Bawden. *User-oriented evaluation of information systems and services*, 49. Brookfield, VT: Gower, 1990.

Davis, Charles H. and Deborah Shaw. 1989. Comparison of retrieval system interfaces using an objective measure of screen design effectiveness. *Library and Information Science Research* 11: 325-334.

Debons, A. and K. L. Montgomery. 1974. Design and evaluation of information systems, in *Annual Review of Information Science and Technology* 9, ed. C. A. Cuadra, A. W. Luke and J. L. Harris. Washington, D.C.: American Society for Information Science. Quoted in Rowena Weiss Swanson. Performing evaluation studies in information science. *Key papers in design and evaluation of information systems*, ed. Donald W. King, 58-74. White Plains, NY: Knowledge Industry Publications, 1978. Reprinted from *Journal of the American Society for Information Science* May-June 1975.

Dervin, Brenda, and Nilan, Michael. 1986. Information Needs and Uses. In *Annual Review of Science and Technology* 21:3-33.

Ellis, David. 1984. The effectiveness of information retrieval systems: the need for improved explanatory frameworks. *Social Science Information Studies* 4:261-272.

_____. 1989. A behavioural approach to information retrieval system design. *Journal of Documentation* 45 (September): 171-212.

_____. 1992. The physical and cognitive paradigms in information retrieval research. *Journal of Documentation* 48 (March): 45-64.

_____. 1993. Modeling the information-seeking patterns of academic researchers: a grounded theory approach. *Library Quarterly* 63 (4) : 469-486

Ellis, David, Nigel Ford and Frances Wood. 1993. Hypertext and learning styles. *The Electronic Library* 11 (February): 13-17.

Garvey, William D. 1979. The dynamic scientific-information user. Chap. in *Communication: The essence of science*. Oxford: Pergamon.

Green, Andrew. 1990. What do we mean by user needs?. *British Journal of Academic Librarianship* 5(2): 65-78.

Hert, Carol A., and Michael S. Nilan. 1991. User-based information retrieval system interface evaluation: An examination of an on-line public access catalog. In *ASIS '91: Proceedings of the 54th ASIS annual meeting Washington, D.C. October 27-31, 1991*, Vol. 28. Medford, NJ: Learned Information, 170-177.

Large, Andrew, Jamshid Beheshti, Alain Breuleux, and Andre Renaud. 1994. A comparison of information retrieval from print and CD-ROM versions of an encyclopedia by elementary school students. *Information Processing and Management* 30(4): 499-513.

Marchionini, Gary and Peter Liebscher. 1991. Performance in electronic encyclopedias: implications for adaptive systems. In *ASIS '91: Proceedings of the 54th ASIS annual meeting Washington, D.C. October 27-31, 1991*, Vol. 28. Medford, NJ: Learned Information, 39-48.

Pandit, Idrisa. 1992. Informal communication in the humanities: A qualitative inquiry. Ph.D Thesis, University of Illinois at Urbana-Champaign.

Patton, M. Q. 1986. *Utilization-focused evaluation* (2nd ed.). Beverley Hills: Sage.

Roberts, Norman. 1982. A search for information man. *Social Science Information Studies* 2: 93-104.

Rouse, William B. 1987. On meaningful menus for measurements: Disentangling evaluative issues in system design. *Information Processing & Management* 23(6): 593-604.

Rogers, A. Robert. 1985. The visible college. Edited by Sinikka Koskiala. *Journal of Education for Library and Information Science* 26 (Summer): 54-55.

Sheehy, Eugene P, ed. 1986. *Guide to Reference Books*. 10th ed. Chicago and London: American Library Association.

Stoan, Stephen K. 1991. Research and information retrieval among academic researchers: Implications for library instruction. *Library Trends* 39 (Winter): 238-57.

Swanson, Rowena Weiss. 1978. Performing evaluation studies in information science. *Key papers in design and evaluation of information systems*, ed. Donald W. King, 58-74. White Plains, NY: Knowledge Industry Publications. Reprinted from the *Journal of the American Society for Information Science* May-June 1975.

Watson-Boone, Rebecca. 1994. The information needs and habits of humanities scholars. *RQ* 34 (Winter): 203-216.

Wilson, T.D. 1981. On user studies and user needs. *Journal of Documentation* 37 (March): 3-15.

Woodsmall, Rose Marie, and Elliot R. Siegal. 1988. Reconciling design philosophy and user expectations. In *Design and evaluation of computer/human interfaces: Issues for librarians and information scientists: Papers presented at the 1988 Clinic on Library Applications of Data Processing, April 17-19, 1988*. N.p.: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 163-182.